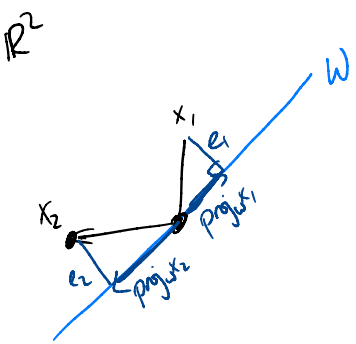
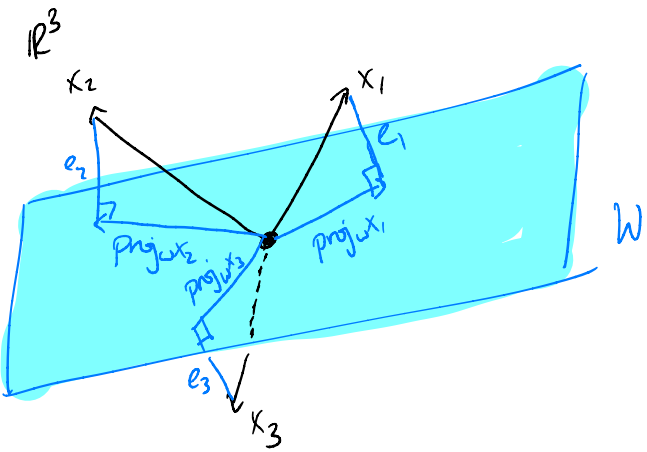
Best Fit Subspaces

Def: Given vectors $x_1, \dots, x_m \in \mathbb{R}^n$, the best-fit k-dimensional subspace W of $x_1, \dots, x_m$ is the subspace of $\mathbb{R}^n$ that minimizes $\sum_{i=1}^{m} \|e_i\|^2$, where $e_i = x_i - \text{proj}_W x_i$





Earlier in the semester, we used least squares to find lines of best fit. This process minimized the sum of squared vertical distances between the data and line.

Here, we minimize the sum of squared perpendicular distances. (We will apply this soon in Principle Component Analysis)

Note: minimizing $\sum_{i=1}^{m} \|e_i\|^2$ is the same as maximizing $\sum_{i=1}^{m} \|\text{proj}_W x_i\|^2$ since, by Pythagorean thm,

$$\sum_{i=1}^{m} \|e_i\|^2 + \sum_{i=1}^{m} \|\text{proj}_W x_i\|^2 = \sum_{i=1}^{m} \|x_i\|^2, \text{ which is a fixed quantity.}$$

We can do this with SVD!

Theorem: Let $A = \begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix}$ and let $A = [u_1 \cdots u_m] \Sigma [v_1 \cdots v_n]^T$ be an SVD of A. Then the best fit k-dim'l subspace of $x_1, \ldots, x_m$ is $W = \text{Span}\{v_1, \ldots, v_k\}$. (First k singular vectors.)

First, an important lemma:

Lemma: Let $A = U\Sigma V^T$ be an SVD where $V = [v_1 \cdots v_n]$. Then $\|Av_1\| \geq \|Aw\|$ $\forall$ unit vectors $w \in \mathbb{R}^n$ (i.e. v_1 maximizes $\|Aw\|$ when w is unit)

Proof: Let w be a unit vector. Then

$$\begin{aligned} \|Aw\|^2 = (Aw)\cdot(Aw) &= (Aw)^T(Aw) \\ &= w^T A^T A w \\ &= w^T (U\Sigma V^T)^T (U\Sigma V^T) w \\ &= w^T V \Sigma^T U^T U \Sigma V^T w \\ &= w^T V \Sigma^T \Sigma V^T w \quad \text{since } U^T U = I \\ &= (V^T w)^T \Sigma^T \Sigma (V^T w) \end{aligned}$$

Notice $\Sigma^T\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}^T \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \ddots & & \\ & & \lambda_r & \\ 0 & & & 0 \end{bmatrix}$

is a diagonal matrix whose diagonal entries are eigenvalues of A^TA ($\lambda_i = \sigma_i^2$), which satisfy $(A^TA)v_i = \lambda_i v_i$ $\forall i$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Let $V^T w = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Then

$$\begin{aligned}
(V^T w)^T \Sigma^T \Sigma (V^T w) &= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \\
&\le \lambda_1 y_1^2 + \cdots + \lambda_1 y_n^2 \\
&= \lambda_1 (y_1^2 + \cdots + y_n^2) \\
&= \lambda_1 \|Vw\|^2 \\
&= \lambda_1 \|w\|^2 \qquad \text{since } V \text{ is orthogonal} \\
&= \lambda_1 \qquad \text{since } \|w\| = 1.
\end{aligned}$$

Hence $\|Aw\| \le \lambda_1$, where w is unit. The maximum value of λ_1 is achieved when $w = v_1$:

$$\|Av_1\|^2 = (V^T v_1)^T \Sigma^T \Sigma (V^T v_1) = e_1^T \Sigma^T \Sigma e_1 = \lambda_1 \quad \checkmark$$

Hence v_1 maximizes $\|Aw\|$ (and has max value $\sqrt{\lambda_1} = \sigma_1$)

We can use the lemma to prove the theorem.

proof for $k=1$: the general case is similar, but more complicated

We need to find a unit vector u such that $\sum_{i=1}^{n} \|\text{proj}_W x_i\|^2$ is maximized, where $W = \text{Span}\{u\}$

We have $\sum_{i=1}^{m} \|\text{proj}_W x_i\|^2 = \sum_{i=1}^{m} \|(u \cdot x_i) u\|^2$ since $u \cdot u = 1$

$$= \sum_{i=1}^{m} (u \cdot x_i)^2 \|u\|^2$$

$$= \sum_{i=1}^{m} (u \cdot x_i)^2 \quad \text{since } \|u\| = 1$$

$$= \left\| \begin{bmatrix} u \cdot x_1 \\ \vdots \\ u \cdot x_m \end{bmatrix} \right\|^2$$

$$= \|Au\|^2$$

By the lemma, $\|Au\|$ is maximized when $u = v_1$.
Hence v_1 maximizes $\sum_{i=1}^{n} \|\text{proj}_w x_i\|^2$.
Hence $\text{Span}\{v_1\}$ is the best fit 1-dimil subspace ■

Ex: Find the 1-dimensional subspace of best fit of the data $x_1 = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$, $x_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $x_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $x_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Let $A = \begin{bmatrix} 1 & 0 & -2 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. By the theorem, the best fit 1-dimil sbsp is spanned by the first singular vector in V.
This is an eigenvector corresponding to the largest eigenvalue of A^TA. Instead of finding SVD, let's find this space directly by finding such an eigenvector

$A^TA = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}$ has eigenvalues $\lambda = 3, 2, 7$

and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an evector corresponding to $\lambda = 7$.

Hence the best fit 1-dimil subspace is $\text{Span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$